Rentomojo Node.js Assignment

Problem Statement:

Recursively crawl popular blogging website https://medium.com using Node.js and harvest all possible hyperlinks that belong to **medium.com** and store them in a database of your choice

What do you need to store ?

1. Every unique URL you encountered.
2. The total reference count of every URL.
3. A complete unique list of parameters associated with this URL.

Things you should keep in mind:

1. Your solution needs to be **asynchronous** in nature.
2. Maintain a **concurrency of 5 requests** at all times, do not end up getting blocked.
3. If you are using **request.js**, you are not allowed to use its connection pool.
4. You are not allowed to use any external scraping or **async** library.
5. Refrain from using **throttled-request** package to limit concurrency.

Things that we love: (Highly encouraged, but not required)

1. A well baked README file
2. Project setup with a simple command
3. A concise project structure with configurations
4. Good commit history with meaningful and atomic commits
5. A dockerized solution

Expected behavior of your solution

Assume your recursive scraper parses 4 URLs:

1. https://medium.com/some/thing
2. https://medium.com/some/thing?param1=abc
3. https://medium/com/some/thing?param2=xyz
4. https://medium/com/some/thing?param1=def&param3=xxx

Your chosen database must contain the URL https://medium.com/some/thing with the a **reference count** of 4. A unique list of parameters containing **param1,param2,param3.** Don't worry about parameter values.

Submission

It is mandatory to submit the assignment in a git repo. You are encouraged to use GitHub, BitBucket or GitLab.